



Approche géométrique et classification pour la reconnaissance de visage

Bertrand Leroy, Ahlame Chouakria, Isabelle Herlin, Edwin Diday

► To cite this version:

Bertrand Leroy, Ahlame Chouakria, Isabelle Herlin, Edwin Diday. Approche géométrique et classification pour la reconnaissance de visage. Congrès de Reconnaissance des Formes et Intelligence Artificielle, Jan 1996, Rennes, France. pp.548-557. inria-00532693

HAL Id: inria-00532693

<https://inria.hal.science/inria-00532693>

Submitted on 20 Apr 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Approche géométrique et classification pour la reconnaissance de visage

Bertrand LEROY, Ahlame CHOUAKRIA, Isabelle L. HERLIN et Edwin DIDAY

INRIA

Domaine de Voluceau, Rocquencourt

B.P. 105, 78153 Le Chesnay Cedex, France.

Tél : (1) 39 63 54 79

Email: Bertrand.Leroy@inria.fr, Ahlame.Chouakria@inria.fr

Isabelle.Herlin@inria.fr, Edwin.Diday@inria.fr

Résumé

Abstract

Nous présentons dans cet article un système d'identification de visage fondé sur l'utilisation des propriétés géométriques. Ce système se décompose en trois phases successives de traitement : une localisation d'éléments caractéristiques du visage, l'extraction de leur frontière par une méthode de contour actif et une analyse factorielle des données ainsi extraites. La localisation des régions contenant les éléments significatifs du visage, tels que la bouche, les yeux et le nez, est effectuée à l'aide d'opérateurs de morphologie mathématique. Afin d'extraire ensuite la frontière de ces structures, nous employons une méthode de contour actif dédiée. Cette méthode est fondée sur l'utilisation de descripteurs de Fourier, ce qui permet d'adapter simplement le modèle à la morphologie de chacun des objets du visage. Les contours, ainsi obtenus sur le visage à identifier, sont ensuite mis en correspondance avec ceux obtenus sur les visages d'une base d'apprentissage. À cet effet, nous proposons une nouvelle méthode d'analyse en composantes principales étendue à des données de type intervalle ; cette méthode s'inscrit dans le cadre de l'analyse de données symboliques.

In this paper we present a system for face recognition based on a geometrical approach. This system is made of three main processing steps: salient face features are first located; their boundaries are then obtained using an active contour model and finally we apply a factorial analysis on these data in order to recognize the face. The localization of the salient features, such as nose, mouth and eyes is obtained by using morphological operators. In order to extract the boundary of these elements, we use a specific active contour method. This model is based on Fourier descriptors and is able to incorporate information about the global shape of each object. Finally, these boundaries, extracted from the face to recognize, are matched with the boundaries coming from a learning data base. To achieve this task, we introduce a new principal component analysis scheme extended to interval data in the framework of symbolic data analysis.

Mots clés : reconnaissance de visage, contour actif, contour spatio-temporel, analyse factorielle, analyse de données symboliques.

Keywords: face recognition, active contour, space-time contour, factorial analysis, symbolic data analysis.

1 Introduction

Les travaux portant sur la reconnaissance du visage, à partir de vues frontales 2D prises dans des conditions d'illumination constantes, peuvent être répertoriés en deux catégories distinctes suivant qu'ils se rapportent à une approche géométrique ou à une approche pixel. L'approche géométrique consiste à extraire la position relative des différents éléments du visage (tels que le nez, la bouche et les yeux) afin de restreindre l'espace des données initiales. Différentes méthodes statistiques, inspirées des algorithmes classiques de reconnaissance de formes, sont alors utilisées pour discriminer les visages à partir de ces mesures [10] [7] [4].

L'approche pixel est une approche globale qui nécessite peu de connaissance préalable sur la structure du visage et se base principalement sur l'information pixel. Certains travaux s'inspirent de méthodes de traitement d'image liées à la mise en correspondance de régions ; les mesures utilisées pour identifier les visages sont alors obtenues par des calculs de corrélation entre les régions. D'autres méthodes utilisent une analyse en composantes principales réalisée à partir de l'information pixel, pour rechercher une base de représentation permettant de discriminer les visages [1] [2] [15].

Le système de reconnaissance de visage, que nous présentons, s'inscrit dans le cadre du projet AMIBE, un des projets de recherche du PRC Communication Homme-Machine, qui se propose d'expérimenter une interface multimodale

homme-machine intégrant le son et l'image pour un nombre limité d'utilisateurs. L'utilisateur d'une telle interface est d'abord identifié par une carte magnétique personnelle puis, tout au long de la transaction, son identité est vérifiée en utilisant les modes parole et image. Parallèlement, la communication entre l'interface et l'utilisateur est gérée par un module de reconnaissance de la parole visuelle et auditive. Les données utilisées sont donc des séquences vidéo de locuteurs filmés de face avec un éclairage fixe.

Afin de permettre l'identification de visage tout en fournissant au module de lecture labiale du système AMIBE des informations sur le mouvement de lèvres, nous proposons un système de reconnaissance de visage fondé sur une utilisation intensive de la géométrie du visage. Dans la section 2, nous présentons une méthode permettant de localiser les éléments significatifs du visage tels que les yeux, le nez et la bouche. Dans la section 3, nous proposons un modèle spécifique de contours actifs à base de descripteurs de Fourier permettant d'extraire la frontière de ces éléments. Ce modèle paramétrique offre l'avantage de pouvoir s'adapter aisément à la géométrie de l'objet étudié. Enfin, dans la section 4 nous présentons un nouveau type d'analyse en composantes principales adaptée à des données intervalles qui permet d'obtenir, à partir des informations extraites d'une base d'apprentissage, une représentation pertinente et compacte du visage. L'identification d'un visage inconnu consiste alors à projeter les caractéristiques mesurées sur ce visage dans la base des composantes principales et à utiliser des mesures de similarité afin de déterminer le visage le plus proche dans la base d'apprentissage.

2 Localisation des éléments caractéristiques du visage

Dans le cadre de l'application AMIBE, l'acquisition des images de visage se fait sous un éclairage fixe et la prise de vue est frontale. Le processus de localisation des éléments du visage est fondé sur l'utilisation de cette particularité d'acquisition. Il s'agit en effet de structures peu mobiles pouvant subir des déformations et qui présentent un contraste important par rapport à leur environnement spatial.

La première tâche définie consiste à localiser les yeux. Le nez et la bouche seront ensuite recherchés en utilisant la géométrie du visage.

L'existence d'une tache de réflexion spéculaire, due aux sources lumineuses frontales sur l'iris, nous permet de localiser l'œil en utilisant un opérateur morphologique de type "pic", qui a pour effet de faire ressortir les maxima locaux de la fonction d'intensité de niveau de gris (voir figure 1). Afin de localiser ces taches de réflexion spéculaire, on utilise un élément structurant carré de 3 pixels de côté ; la taille des images étant de 704 par 576 pixels. Après application d'un seuillage sur l'image pic, la position des iris est déterminée en balayant l'image de haut en bas à la recherche des deux premières régions d'intensité non nulle.

Pour localiser le nez et la bouche, nous utilisons ensuite l'information spatio-temporelle de façon à caractériser les points ayant enregistré les plus fortes variations d'intensité au cours de la séquence. À cet effet, nous réalisons en chaque point la sommation de la norme du gradient spatio-temporel (voir figure 2). L'image I_{som} ainsi obtenue vérifie :

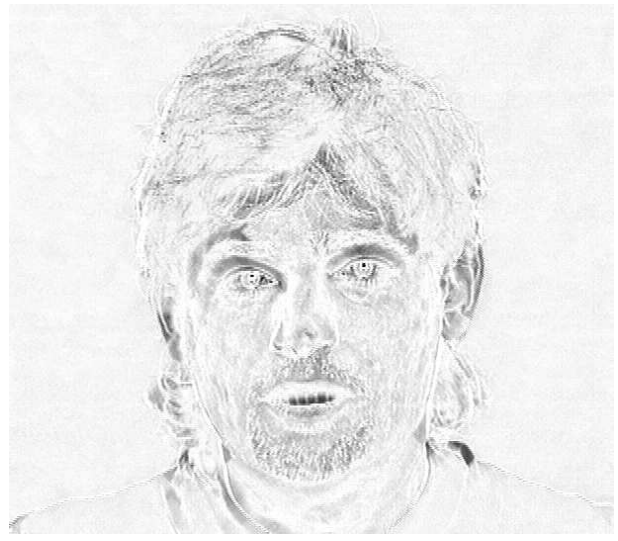


FIG. 1 - Inverse du résultat obtenu par application de l'opérateur morphologique de type "pic".

$$I_{som} = \sum_{i=1}^N \sqrt{\left(\frac{\partial I}{\partial x}\right)^2 + \left(\frac{\partial I}{\partial y}\right)^2 + \mu \left(\frac{\partial I}{\partial t}\right)^2}, \quad (1)$$

où $I(x, y, t)$ représente l'intensité de niveau de gris ; x, y les composantes spatiales ; t la composante temporelle et N le nombre d'images de la séquence. μ est un terme de normalisation dépendant de l'échantillonnage temporel de la séquence et du mouvement du visage.

L'image I_{som} permet de mettre en valeur les régions de faible contraste mais ayant une grande mobilité et les régions peu mobiles et à fort contraste.

La région de la bouche étant une région très mobile, il est intéressant de considérer la composante temporelle du gradient pour la localiser. La partie inférieure du nez est une région à fort gradient spatial car son relief crée des ombres sur le visage. Cette région est plutôt statique mais, si le visage bouge un peu lors de la séquence, la composante temporelle du gradient sera également importante. Il est donc préférable de prendre en compte à la fois les composantes temporelle et spatiales pour caractériser ces structures.

La localisation du nez et de la bouche sur l'image I_{som} consiste à rechercher, le long de la médiatrice du segment passant par les iris, les deux régions ayant les plus fortes valeurs sur l'image I_{som} . On obtient ainsi avec précision la position de la partie inférieure du nez. Le point obtenu, par cette méthode, pour localiser la bouche dépend en fait de l'activité labiale de la personne au cours de la séquence. Si cette activité est importante, le point est généralement à l'intérieur de la bouche sinon c'est un point situé sur le bord des lèvres qui est retenu.

Afin d'accélérer le traitement, les images utilisées pour la localisation sont les trente premières de la séquence vidéo, ce qui correspond à une durée d'acquisition proche de la seconde. Utiliser une séquence plus courte aurait pour effet de diminuer la précision des résultats car la norme de la composante temporelle du gradient est assez faible.

3 Extraction des éléments caractéristiques du visage



FIG. 2 - Inverse du résultats sur deux visages différents du calcul de la valeur cumulée de la norme du gradient spatio-temporel sur une séquence d'images.



FIG. 3 - Résultat de la localisation des yeux, du nez et de la bouche.

L'utilisation de modèles de contour actif, afin d'extraire les éléments du visage, suppose que l'on dispose d'un modèle capable de détecter avec précision les contours d'objets pouvant subir de fortes déformations (pour permettre la caractérisation spatio-temporelle de la bouche) et possédant des coins (pour permettre l'étude des yeux et de la bouche). Les modèles de contour actif génériques tels que les "snakes" [3] [11] ne permettent pas de prendre en compte la forme spécifique des objets. Plus exactement, les contraintes géométriques appliquées à ces modèles apparaissent sous la forme de paramètres énergétiques et ne sont pas intégrées au modèle.

Pour ces raisons, des études spécifiques ont été réalisées afin de définir des modèles adaptés à l'extraction des contours d'éléments du visage ; ces modèles sont généralement des modèles déformables paramétriques.

Yuille, Cohen et Hallinan [16] ont proposé un ensemble de modèles déformables spécifiques à chaque élément du visage. Ainsi le modèle utilisé pour la caractérisation de l'œil est constitué de 11 paramètres (le centre de l'œil, les foyers des demi-ellipses formant les contours supérieur et inférieur, l'orientation, le rayon de l'iris, ...).

Craw, Tock et Bennett [4] ont décrit un modèle déformable global permettant d'extraire près de 40 points caractéristiques sur une image de visage. Ce modèle est constitué d'un ensemble de modules adaptés au traitement de chacun des éléments spécifiques du visage et d'une structure globale contrôlant la position géométrique des différents modules les uns par rapport aux autres. Chaque module est gouverné par un ensemble de contraintes statistiques qui lui est spécifique. Cette méthode donne des résultats intéressants concernant la localisation des éléments du visage ; cependant, elle ne permet pas d'extraire avec précision les contours de ces éléments et nécessite un temps de calcul important.

Nous présentons dans cette section un modèle paramétrique de contour actif à base de descripteurs de Fourier. L'objectif est de disposer d'un modèle simple et suffisamment général pour pouvoir extraire la bouche, les yeux et le nez tout en tenant compte de la géométrie propre de chacun de ces objets. L'utilisation de descripteurs de Fourier est particulièrement adéquate dans le cas de la bouche et des yeux puisque la forme de ces objets est proche de celle d'une ellipse. Le nez est, quant à lui, modélisé par une courbe ouverte.

3.1 Courbes elliptiques fermées

Staib et Duncan [14] ont été les premiers à proposer l'utilisation de tels descripteurs pour modéliser des contours actifs. Leur méthode de convergence est une méthode stochastique bayésienne. Pour déterminer la distribution de probabilités utilisée pour la convergence, il est nécessaire d'avoir une connaissance précise de la géométrie et de la position de l'objet recherché. En effet, chaque paramètre suit une distribution de probabilités selon une loi normale centrée sur la valeur initiale de ce paramètre. Dans le cadre du système AMIBE, il n'est toutefois pas possible de connaître a priori la forme de chaque objet à un instant origine et leur localisation n'est pas suffisamment précise pour permettre l'utilisation d'une telle méthode de convergence. Il est donc préférable d'utiliser une méthode variationnelle pour rechercher la solution du modèle.

Une courbe elliptique fermée est une courbe paramétrée dont la position de chaque point est donnée par :

$$\begin{aligned} v(\theta) &= \begin{pmatrix} x(\theta) \\ y(\theta) \end{pmatrix} \\ &= \sum_{k=0}^N A_k \begin{pmatrix} \cos(\theta k) \\ \sin(\theta k) \end{pmatrix}, \end{aligned} \quad (2)$$

où A_k est une matrice 2×2 ; N le nombre d'harmoniques utilisées pour décrire cette courbe et θ un paramètre angulaire.

Afin d'extraire les contours d'un objet, nous recherchons un minimum local de la fonctionnelle d'énergie :

$$\begin{aligned} E(v) &= \int_0^{2\pi} P(v(\theta)) + \lambda \frac{\partial v(\theta)^2}{\partial \theta} d\theta, \\ \text{où } P &= -|\nabla I * G|^2 \text{ et } \lambda \in \mathbb{R}^+. \end{aligned} \quad (3)$$

Le premier terme de l'énergie est lié à l'image traitée. P correspond au carré de la norme du gradient calculé dans l'image I et convolué par un filtre de lissage G . Le second terme de l'énergie est un terme d'élasticité lié à la tension de la courbe. Le paramètre λ est un poids associé à la contrainte d'élasticité et il permet d'effectuer une pondération entre les forces d'élasticité et les forces issues du potentiel image.

Le gradient de $E(v)$ est un vecteur de dimension $4 \times N$ dont les composantes sont les dérivées partielles de E par rapport à chacun des quatre éléments des N matrices A_k :

$$\frac{\partial E}{\partial p} = \int_0^{2\pi} \frac{\partial P(v(\theta))}{\partial v(\theta)} \cdot \frac{\partial v(\theta)}{\partial p} + 2\lambda \frac{\partial v(\theta)}{\partial \theta} \frac{\partial v(\theta)}{\partial \theta} d\theta, \quad (4)$$

où p est un élément de A_k .

La recherche de la courbe v^* , réalisant un minimum local de E , est effectuée à l'aide d'une méthode de minimisation de type Newton. Afin d'assurer une meilleure convergence, la minimisation est d'abord effectuée sur les paramètres de basses fréquences puis sur les paramètres de fréquences plus élevées.

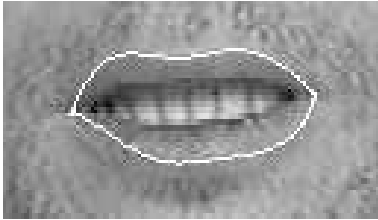


FIG. 4 - Utilisation de courbes elliptiques fermées à cinq harmoniques pour modéliser le contour extérieur de la bouche.

Plusieurs tests ont été effectués avec différents potentiels et différents nombres d'harmoniques. Ces tests montrent que le fait d'augmenter le nombre de modes (i.e. harmoniques) utilisés pour décrire la courbe n'améliore pas forcément les résultats. Contrairement aux "snakes", il n'est pas nécessaire d'ajouter à la fonctionnelle d'énergie un caractère régularisant puisque le nombre d'harmoniques décrivant le modèle détermine à lui seul les caractéristiques de régularité de la courbe. Ainsi, si l'on relâche les contraintes de régularité en utilisant des modes de fréquence élevée, on constate l'apparition d'oscillations parasites.

3.2 Courbes elliptiques ouvertes connexes

L'utilisation de courbes elliptiques fermées est problématique lorsque le contour de l'objet recherché possède des points fortement irréguliers au sein de portions régulières (c'est le cas des yeux et de la bouche dont les contours contiennent des coins). Ainsi, pour pouvoir modéliser des coins, il est nécessaire d'utiliser des paramètres de fréquences élevées, ce qui a pour effet de créer des oscillations sur les portions de courbes devant être régulières. Afin de remédier à ce problème, on utilise des modèles elliptiques ouverts. Les courbes ouvertes sont généralement utilisées pour représenter des contours d'objets n'ayant pas de frontière complète ou pour décrire des segments de courbe comme les lettres [13]. Dans le cas présent, un modèle formé de deux courbes elliptiques ouvertes et connexes permet de représenter de façon fiable le contour de la bouche ou des yeux. Le contour du nez est par contre décrit par une seule courbe ouverte.

Soit v_N un modèle formé de deux courbes ouvertes, v_A et v_B , jointes en leurs extrémités et défini par :

$$v(\theta) = \begin{cases} v_A(\theta) & \text{si } 0 \leq \theta < \pi \\ v_B(\theta) & \text{si } \pi \leq \theta < 2\pi \end{cases} \quad (5)$$

v_A et v_B respectent de plus les conditions :

$$v_A(0) = v_B(2\pi) \text{ et } v_A(\pi) = v_B(\pi). \quad (6)$$

La fonctionnelle d'énergie associée au modèle devient :

$$\begin{aligned} E(v) &= \int_0^\pi P(v_A(\theta)) + \lambda \left(\frac{\partial v_A(\theta)}{\partial \theta} \right)^2 d\theta \\ &+ \int_\pi^{2\pi} P(v_B(\theta)) + \lambda \left(\frac{\partial v_B(\theta)}{\partial \theta} \right)^2 d\theta. \end{aligned} \quad (7)$$

L'utilisation de descripteurs de Fourier dans le cas de courbes ouvertes nécessite la définition d'un nouveau modèle. En effet, afin d'être cohérent avec la formulation de l'énergie définie par l'équation (7), il convient de choisir une base différente de celle utilisée en (2). On utilise donc une base de cosinus pour décrire $x(\theta)$ et de sinus pour décrire $y(\theta)$. Ces bases sont orthogonales sur $[0, \pi]$ et il suffit d'une seule harmonique pour décrire une demi-ellipse lorsque θ varie de 0 à π . L'ensemble des courbes décrites par ces bases est représenté par :

$$v_A(\theta) = \begin{pmatrix} a_0 \\ b_0 \end{pmatrix} + R_\phi \sum_{k=1}^N \begin{pmatrix} a_k \cos(\theta k) \\ b_k \sin(\theta k) \end{pmatrix}, \quad (8)$$

où a_k et b_k sont les paramètres du modèle, N le nombre d'harmoniques utilisées pour décrire la courbe et R_ϕ la matrice de rotation d'angle ϕ par rapport à l'origine. Le déphasage existant entre les descripteurs utilisés pour $x(\theta)$ et $y(\theta)$ permet de décrire une demi-ellipse sans être obligé d'utiliser un nombre important d'harmoniques.

Dans la pratique, les contours de la bouche sont modélisés par une courbe à une harmonique pour la partie inférieure et cinq pour la partie supérieure. De même, on utilise une harmonique pour la partie inférieure des yeux et trois pour la partie supérieure.

Soit v_A une courbe définie dans une base elliptique de Fourier et décrite par l'équation (8), la courbe v_B décrite par une seule harmonique et répondant aux conditions données en (6) est représentée par :

$$v_B(\theta) = \begin{pmatrix} \sum_{i=0}^{N/2} a_{2i} \\ b_0 \end{pmatrix} \quad (9)$$

$$+ R_\phi \begin{pmatrix} \left(\sum_{i=0}^{N/2} a_{2i+1} \right) \cos(\theta) \\ c \sin(\theta) \end{pmatrix}. \quad (10)$$

Les gradients de l'énergie par rapport aux différents paramètres vérifient :

$$\frac{\partial E(v_N)}{\partial a_{2k+1}} = \int_0^\pi \frac{\partial P(v_A)}{\partial x_a} \cos(k\theta) d\theta + \pi \sum_{i=1}^{N/2} a_{2i+1}$$

$$+ \int_\pi^{2\pi} \frac{\partial P(v_B)}{\partial x_b} \cos(k\theta) d\theta + \pi k^2 a_k$$

$$\frac{\partial E(v_N)}{\partial a_{2k}} = \int_0^\pi \frac{\partial P(v_A)}{\partial x_A} \cos(k\theta) d\theta + \pi k^2 a_k$$

$$+ \int_\pi^{2\pi} \frac{\partial P(v_B)}{\partial x_B} \cos(k\theta) d\theta$$

$$\frac{\partial E(v_N)}{\partial b_k} = \begin{cases} \int_0^\pi \frac{\partial P(v_A)}{\partial y_A} d\theta + \pi k^2 a_k \\ + \int_\pi^{2\pi} \frac{\partial P(v_B)}{\partial Y_b} d\theta & \text{si } k = 0 \\ \int_0^\pi \frac{\partial P(v_A)}{\partial Y_a} \sin(k\theta) d\theta \\ + \pi k^2 a_k & \text{sinon} \end{cases}$$

$$\frac{\partial E(v_N)}{\partial c} = \int_\pi^{2\pi} \frac{\partial P(v_B)}{\partial y_B} \sin(\theta) d\theta + \pi c$$

Un examen des dérivées partielles de $E(v)$ montre que les dérivées partielles par rapport aux paramètres a_k n'ont pas le même poids que celles calculées par rapport aux paramètres b_k et c . Ainsi, les forces exercées sur l'axe horizontal sont deux fois supérieures aux forces verticales, toute chose égale par ailleurs. Pour améliorer la convergence, on pondère donc les dérivées de $E(v)$ par rapport aux paramètres a_k par un facteur de $1/2$. Les résultats obtenus par une minimisation de type Newton, en utilisant cette pondération, sont des minima locaux de $E(v)$; il n'est pas nécessaire de modifier la fonctionnelle d'énergie.

Si l'on se place dans le cas particulier d'un potentiel image nul ($P(x, y) = 0$), le résultat obtenu en minimisant la fonctionnelle d'énergie à partir d'une courbe initiale représentée par une ellipse est identique avec ou sans pondération : c'est une courbe de longueur nulle. Cependant, les courbes obtenues à chaque itération de la descente du gradient sont différentes. Sans pondération, le grand axe se réduit plus rapidement que le petit axe et la courbe devient un cercle puis tend vers une ellipse dégénérée ayant un axe horizontal nul. Avec pondération, le rapport du grand axe au petit axe reste constant. Ainsi, la force de rétraction appliquée à la courbe

n'a pas de caractère régularisant. Il est alors possible d'influencer la géométrie de la courbe obtenue au minimum local de la fonctionnelle d'énergie par le biais de contraintes sur l'initialisation.

3.3 Obtention des éléments du visage

Les algorithmes utilisés pour détecter les contours des éléments du visage sont de type pyramidal. Ainsi, partant de deux demi-ellipses comme courbe initiale entourant l'objet étudié, on recherche une courbe v_1 décrite par une seule harmonique. Le résultat obtenu est alors utilisé comme initialisation d'un modèle v_2 , à deux harmoniques, jusqu'à l'obtention d'une courbe v_N , à N harmoniques, décrivant les contours de l'élément recherché.



FIG. 5 - Extraction du contour des yeux. Initialisation et résultats obtenus avec une, puis trois harmoniques pour le contour supérieur. Le contour inférieur est décrit par une seule harmonique.



FIG. 6 - Extraction du contour de la bouche. Initialisation et résultats obtenus avec une, puis cinq harmoniques pour le contour supérieur. Le contour inférieur est décrit par une seule harmonique.

La force d'élasticité qui pousse le contour à se rétracter vers l'objet joue un rôle déterminant lors de la première étape. Dans les étapes suivantes, le nombre d'harmoniques augmente et cette force exerce uniquement un effet de rappel sur la courbe. Cette méthode permet de converger rapidement et sûrement vers une approximation du contour de l'objet.

Les figures 6 et 7 donnent un aperçu des résultats obtenus à différents moments du processus de recherche des contours du nez et de la bouche. L'expérience montre qu'une courbe elliptique ouverte à cinq harmoniques permet de modéliser les contours du nez avec une précision satisfaisante. La bouche est décrite par un modèle de deux courbes ouvertes connexes utilisant cinq harmoniques pour la partie supérieure et une harmonique pour la partie inférieure.

Le voisinage des yeux forme une région riche en contours (l'iris, les paupières, les sourcils, ...) et il est nécessaire d'utiliser une initialisation précise afin de détecter les yeux. Cette détection est effectuée en deux temps. Dans un premier temps, la tache de réflexion spéculaire située à l'intérieur de l'iris est localisée et on recherche les contours de l'iris. Le diamètre et la position de l'iris permet alors d'initialiser le modèle de contours actifs de sorte que la courbe ne soit pas attirée par des contours externes aux yeux. La figure 5 montre les résultats obtenus à l'issue des différentes étapes de l'algorithme.



FIG. 7 - Extraction du contour du nez. Initialisation et résultats obtenus avec une, puis cinq harmoniques.

4 Classification

La base de connaissance de notre système de reconnaissance de visage est composée de vecteurs dont chaque composante représente la position d'un élément du visage. L'utilisation de ces informations à des fins d'identification nécessite que l'on définisse une représentation des données qui soit à la fois discriminante et compacte. Nous proposons donc une méthode d'analyse en composantes principales étendue à des données de type intervalle. Cette méthode permet d'estimer les proximités et les recouvrements entre les différentes classes de visages et de déterminer pour un visage particulier et inconnu son degré d'appartenance à chacune des classes de référence.

Afin de valider cette méthode de classification, nous disposons pour chacune des neuf personnes de notre base de données de trois séquences de dix images. Les images de chaque séquence sont extraites d'un court enregistrement vidéo.

4.1 Extension de l'analyse en composantes principales à des données de type intervalle

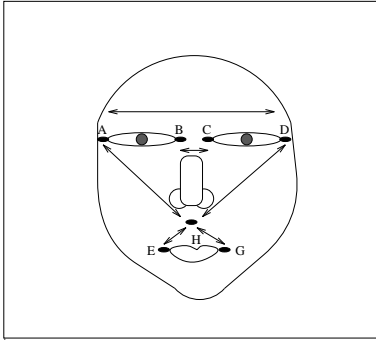


FIG. 8 - Ensemble des points utilisés pour l'identification du visage.

Chaque visage est identifié par des distances calculées entre des points caractéristiques du visage (voir figure 8). Du fait de la mobilité de la tête, les positions des points ainsi que les distances caractérisant le visage, varient lors des différentes prises de vue. Afin de prendre en compte cette variation, un visage sera décrit par des données de type intervalle. On appellera objet visage, un objet décrit par des données complexes (intervalle, ensemble, ...).

Un outil de description des objets complexes et symboliques a été proposé par Diday [5] [6]. Ce formalisme permet d'une part, de décrire des objets par des variables multivaluées (à valeur intervalle, à valeur ensembliste, à valeur probabiliste, ...), et d'autre part de prendre en compte les différents liens et contraintes liant les objets ou les parties composant l'objet. Différents travaux se sont intéressés à

l'extension des méthodes de réduction de dimension et de transformation de variables à des données complexes. Nagabushan [12] a présenté une méthode de réduction à deux dimensions s'appliquant à des objets décrits par des variables à valeurs intervalles. Cette méthode est fondée sur le développement en séries de Taylor. Ichino [8] [9] s'est également intéressé au problème de réduction de dimension, il propose une extension de la méthode d'ACP à des objets décrits par des variables de type intervalle.

La méthode décrite dans cette section, appelée méthode des sommets, permet de projeter les objets visage dans un espace de dimension réduite et d'expliquer chaque groupe d'objets projetés par un nombre réduit de variables de type intervalle. Un avantage important de cette méthode par rapport à celle proposée par Ichino est de fournir une représentation des classes sous forme d'hyper-cube permettant une identification simple et rapide d'un objet.

Soit un tableau de données (d_{ij}) de dimension $(mp \times n)$ contenant les n mesures réalisées sur une séquence de p visages pour chacune des m personnes de la base de données. À partir de ce tableau, on calcule le tableau (d'_{ij}) de dimension $(m \times n)$ où chaque élément est de type intervalle. Ainsi, un élément d'_{ij} de ce tableau regroupe l'information correspondant à la mesure j effectuée sur l'ensemble des p images acquises pour la personne i . Plus précisément, d'_{ij} est l'intervalle défini par la plus petite et la plus grande valeur de la mesure j pour la personne i . L'analyse en composantes principales est effectuée sur ces données de type intervalle et les images de visage seront donc décrites en sortie de l'algorithme par des composantes principales de type intervalle. Ainsi un visage sera représenté, dans le plan factoriel défini par les deux premières composantes principales de type intervalle, par un rectangle dont les longueurs des côtés correspondent à la taille des intervalles de variation de ces composantes principales.

4.2 Méthode de l'analyse

Soit T_i un objet visage décrit par n vecteurs X_1, \dots, X_n . Les composantes de chacun de ces vecteurs représentent les valeurs prises par une variable sur les p différentes images d'un même visage.

$$T_i = \begin{pmatrix} x_1^1 & \cdots & x_n^1 \\ \vdots & \ddots & \vdots \\ x_1^p & \cdots & x_n^p \end{pmatrix} \quad (11)$$

Soit S_i une description sous forme d'intervalles de l'objet visage décrit par T_i :

$$S_i = \begin{pmatrix} [x_{i1}, \overline{x}_{i1}] \\ \vdots \\ [x_{in}, \overline{x}_{in}] \end{pmatrix} \quad (12)$$

où x_{ij} est la plus petite valeur de la variable X_j pour l'objet T_i et \overline{x}_{ij} la plus grande valeur de la variable X_j pour l'objet T_i .

Un objet S_i est donc visualisé, dans l'espace de description, par un hypercube à 2^n sommets. La longueur de chacun des côtés de l'hypercube est donnée par l'étendue de l'intervalle associé à la variable de description. Dans un espace de dimension n , un hypercube peut être décrit par une matrice à 2^n lignes et n colonnes où la $k^{ème}$ ligne correspond aux coordonnées du $k^{ème}$ sommet.

La matrice M_i des sommets de l'hypercube représentant l'objet S_i dans un espace de dimension n est donc définie comme suit :

$$M_i = \begin{pmatrix} \frac{x_{11}}{x_{(i2^n)_1}} & \cdots & \frac{x_{1n}}{x_{(i2^n)_n}} \\ \vdots & \ddots & \vdots \\ \frac{x_{(i2^n)_1}}{x_{(i2^n)_1}} & \cdots & \frac{x_{(i2^n)_n}}{x_{(i2^n)_n}} \end{pmatrix} \quad (13)$$

Lorsque n vaut 2, l'objet S_i :

$$\begin{pmatrix} \frac{x_1}{x_1} & \frac{x_2}{x_2} \\ \frac{x_1}{x_1} & \frac{x_2}{x_2} \end{pmatrix} \quad (14)$$

sera représenté par la matrice suivante :

$$\begin{pmatrix} \frac{x_1}{x_1} & \frac{x_2}{x_2} \\ \frac{x_1}{x_1} & \frac{x_2}{x_2} \end{pmatrix} \quad (15)$$

Il est donc possible de caractériser un objet visage soit par un vecteur de composantes de type intervalle, soit par une matrice de composantes numériques.

Les différentes étapes de la méthode sont :

1. Calculer les vecteurs S_i représentant les objets visage T_i sous forme d'intervalles.
2. Décrire chaque objet S_i par une matrice de données numériques M_i à 2^n lignes et n colonnes.
3. Construire une matrice M possédant $2^n m$ lignes et n colonnes à partir des m matrices M_i associées aux m objets S_i :

$$M = \begin{pmatrix} M_1 \\ \vdots \\ M_m \end{pmatrix} \quad (16)$$

4. Appliquer l'ACP classique à la matrice M de données numériques définie en (16) et (13).
5. Déterminer les composantes principales à valeurs intervalles Y_1^I, \dots, Y_q^I à partir des q premières composantes principales à valeurs numériques Y_1, \dots, Y_q ($q < n$).

Soient :

- L_{S_i} l'ensemble des numéros de lignes dans la matrice M associés à l'objet S_i et
- y_{kj} , la valeur de la $j^{ème}$ composante principale numérique Y_j associée à un sommet k (avec $k \in L_{S_i}$) de l'objet S_i .

La valeur de la $j^{ème}$ composante principale de type intervalle Y_j^I pour l'objet S_i est représentée par :

$$y_{S_i j} = [\underline{y}_{ij}, \overline{y}_{ij}] = [\min_{k \in L_{S_i}} (y_{kj}), \max_{k \in L_{S_i}} (y_{kj})]. \quad (17)$$

4.3 Paramètres d'aide à l'interprétation

Afin de pouvoir évaluer la représentation des visages retenue et de dégager les principales classes ainsi que les éléments caractérisant chaque classe de visage, il convient de définir des mesures de contribution adéquates, prenant en compte la notion d'intervalle.

Contribution relative d'un visage (COR_I) Cette contribution permet de mesurer la qualité de la représentation d'un visage dans le nouvel espace de projection. On obtient la contribution relative $COR_I(S_i, u_j)$ d'un objet visage S_i par rapport à l'axe de direction u_j en calculant le rapport de la somme des inerties des sommets S_i par rapport à l'axe u_j et de la somme des inerties de ces mêmes sommets par rapport au centre de gravité G .

La contribution relative $COR_I(S_i, u_j)$ est donnée par la formule :

$$COR_I(S_i, u_j) = \frac{\sum_{k \in L_{S_i}} p_k \cdot y_{kj}^2}{\sum_{k \in L_{S_i}} p_k \cdot d^2(k, G)} \quad (18)$$

avec

- L_{S_i} l'ensemble des numéros de lignes dans la matrice A associés à l'objet S_i ,
- y_{kj} la coordonnée du sommet k ($k \in L_{S_i}$) de l'objet S_i sur l'axe factoriel de direction u_j ,
- p_k le poids de l'objet S_i
- G le centre de gravité,
- $d(k, G)$ la distance entre i et G .

Contribution absolue d'un visage (CTR_I) On définit la contribution absolue d'un visage S_i à la construction de l'axe de direction u_j comme la somme des contributions absolues des sommets associés. Cette contribution est une mesure de la part de variance apportée par un visage donné. Soit λ_j la valeur propre associée à l'axe factoriel de direction u_j .

$$CTR_I(S_i, u_j) = \frac{\sum_{k \in L_{S_i}} p_k \cdot y_{kj}^2}{\lambda_j} \quad (19)$$

Contribution d'un visage à l'inertie totale (INT_I) La contribution à l'inertie totale de l'objet visage S_i est définie comme la somme des contributions à l'inertie totale des sommets de l'hypercube associé au visage S_i . La contribution à l'inertie totale de l'objet S_i est donnée par la formule (20) :

$$INT_I(S_i) = \frac{\sum_{k \in L_{S_i}} p_k \cdot d^2(k, G)}{INT} \quad (20)$$

où INT représente l'inertie totale.

Afin de valider cette méthode de classification, trois bases de tests sont utilisées. Elles sont issues de la base de données visage qui est constituée d'acquisitions temporelles des visages de neuf personnes de sexe masculin :

- une base de test composée des 10 premières images de 9 séquences vidéo (une séquence vidéo par personne).
- deux bases de test constituées de 10 images correspondant à des plans espacés des enregistrement vidéo. Ces deux bases sont disjointes.

4.3.1 Données numériques

Le tableau des données numériques est composé de 270 lignes et 6 colonnes. Chaque ligne du tableau correspond à une image de visage identifiée par 6 mesures. Ces mesures sont des distances (en nombre de pixels) calculées entre des points particuliers du visage (voir figure 8). Chaque suite de 30 lignes correspond à 30 images de visage d'une même personne. Ces 30 lignes se décomposent elles-mêmes en 3 séries successives de 10 lignes (une série par base de tests).

	PC2	PC5	PC3	PC1	PC6	PC4
AD	0.65	-0.16	0.10	0.64	0.02	-0.37
BC	0.56	0.64	0.29	0.44	-0.04	0.05
AH	-0.35	-0.07	0.33	0.87	-0.10	0.05
DH	0.11	-0.01	-0.30	0.93	0.08	0.17
EH	0.46	-0.18	0.58	-0.38	0.37	0.38
GH	0.76	-0.18	-0.03	-0.36	-0.36	0.36

TAB. 1 - Matrice des corrélations entre les variables x_i et les facteurs principaux.

4.3.2 Données de type intervalle

La transformation des données numériques en données de type intervalle consiste à remplacer chaque série de 10 lignes (une série par classe) du tableau de données numériques par une seule ligne. Les éléments de cette ligne représentent alors les intervalles associés à chacune des 6 mesures. Ces intervalles sont obtenus en prenant la valeur minimale et la valeur maximale de la mesure sur les 10 lignes. Le tableau des intervalles ainsi obtenu est composé de 27 lignes et 6 colonnes.

4.4 Résultats et interprétations

La figure 10 représente la projection des objets visage dans le plan factoriel défini par les deux premières composantes principales.

Les taux d'inertie relatifs aux deux premières composantes principales valent 57, 8% et 20, 35%. Cette méthode fait ressortir huit classes identifiant chacune les visages des personnes FRA, HUS, INC, ISA, KHA, LOT, ROM et une classe regroupant les visages de deux personnes JPL, PHI. Les corrélations entre les variables de départ et les composantes principales indiquent que la variable DH (distance entre le coin extérieur des yeux et le bas du nez), corrélée à 93% avec PC1, représente bien l'axe 1. Les contributions relatives élevées des personnes FRA, INC, ISA, ROM par rapport à l'axe 1 montrent que cet axe distingue particulièrement bien ces personnes.

Le groupe ISA et ROM est caractérisé par des valeurs élevées de DH les opposant directement au groupe INC et FRA qui, lui, est caractérisé par des valeurs faibles des distances DH . La distance GH (distance entre la commissure des lèvres et le milieu du nez) est corrélée à 76% avec l'axe 2. De même, les contributions relatives élevées des personnes HUS, KHA, LOT par rapport à l'axe 2 signifient que ces personnes sont bien représentées par cet axe. De plus, HUS est caractérisé par une distance GH élevée l'opposant au groupe KHA et LOT qui est plutôt caractérisé par des valeurs faibles de GH . On s'abstient, par contre, d'interpréter les positions de JPL et PHI tenant compte des valeurs assez faibles de leurs contributions relatives sur les axes.

Afin de pouvoir discriminer les visages JPL et PHI il est possible de les représenter dans un autre plan factoriel. L'utilisation de la première et de la troisième composante principale (PC1 et PC3) permet d'obtenir une représentation adéquate. La figure 11 montre que la projection des objets visage JPL et PHI dans le plan factoriel défini par PC1 et PC3 fait apparaître deux classes correspondant à chacun de ces deux visages.

Les corrélations répertoriées dans le tableau 1 montrent que la première et la deuxième composante principale peuvent être identifiées à deux variables mesurant, respectivement, la longueur de la partie supérieure du visage et la longueur de la partie inférieure du visage (voir figure 9).

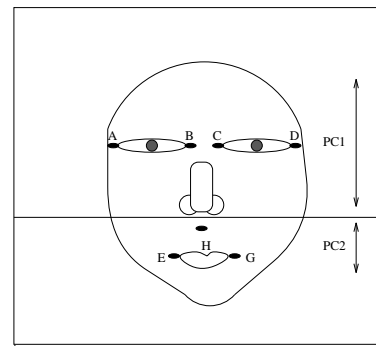


FIG. 9 - Sémantique des composantes principales dans l'identification du visage.

5 Conclusion et perspectives

Nous avons présenté un système permettant de reconnaître les visages à partir d'images 2D. Ce système s'appuie sur l'utilisation de propriétés géométriques du visage.

Pour extraire les points particuliers identifiant un visage, nous proposons un modèle de contour actif paramétrique à base de descripteurs de Fourier. Ce modèle permet de décrire des contours possédant des coins tels que ceux de la bouche ou des yeux en utilisant un faible nombre de paramètres. La fonctionnelle d'énergie associée à ce modèle contient un terme d'élasticité permettant une initialisation relativement éloignée de l'objet recherché. Contrairement à d'autres méthodes, il n'est pas nécessaire d'incorporer à cette énergie un terme de régularité car elle est contenue dans le modèle. Les résultats obtenus par cette méthode permettent d'obtenir une mesure des distances entre les points particuliers du visage.

Afin d'analyser les informations ainsi extraites, nous proposons une méthode originale d'analyse en composantes principales étendue à des données de type intervalles. Cette méthode permet d'obtenir une représentation discriminante des visages tout en tenant compte de la mobilité de certains éléments tels que la bouche.

À partir de cette analyse, il est possible de déterminer aisément le degré d'appartenance d'un visage à une classe en projetant le vecteur caractérisant ce visage dans l'espace des composantes principales. Cette méthode d'identification complétera notre système qui devrait être amélioré en utilisant une méthode pyramidale de classification. Ce type de méthode permet de discriminer les classes ayant des représentations proches dans l'espace des deux premières composantes principales en projetant les groupes de classes voisines dans des espaces définis par des composantes principales moins significatives. Ainsi, lorsqu'un visage est reconnu comme appartenant à un groupe de classes, on utilisera un espace de composantes principales spécifique à ce groupe afin de déterminer à l'intérieur de ce groupe la classe identifiant le mieux le visage.

Une autre perspective de ce travail consistera à étudier une méthode d'analyse factorielle utilisant la forme des contours des éléments caractéristiques du visage plutôt que des distances entre ces éléments.

Remerciements

Nous remercions l'ensemble des partenaires du projet AMIBE pour les discussions fructueuses que nous avons eues :

- ICP - URA 368 - Institut de la Communication Parlée, Grenoble,
- IRT - URA 1399 - Institut de Recherche en Informatique de Toulouse,
- LAFORIA-IBP - URA 1095 - Laboratoire de reconnaissance des FORMES et d'Intelligence Artificielle,
- LIUPPV - Laboratoire d'Informatique d'Avignon,
- LIUM - Laboratoire d'Informatique de l'Université du Mans.

Références

- [1] R. J. Baron. Mechanisms of human facial recognition. *The International Journal of Robotics Research*, 15:137, 1981.
- [2] P. Burt. Algorithms and architecture for smart sensing. In *Proceedings of Image Understanding Workshop*, 1988.
- [3] L. D. Cohen and I. Cohen. A finite element method applied to new active contour models and 3-D reconstruction from cross sections. In *IEEE Proceedings of the International Conference on Computer Vision*, pages 587–591, Osaka, Japan, December 1990.
- [4] I. Craw, D. Tock, and A. Bennett. Finding face features. In *Proceedings of the First European Conference on Computer Vision*, pages 92–96, 1992.
- [5] E. Diday. From data to knowledge, boolean, probabilistic, possibilist and belief objects for symbolic data analysis. In *Tutorial at the fourth Conference of International Federation of Classification Societies*, 1993.
- [6] E. Diday and Y. Kodratoff. *Induction symbolique et numérique partir des données*. Cépadués éditions, 1991.
- [7] M. A. Fischler and R. A. Eschlager. The representation and matching of pictorial structures. *IEEE Transactions on Computers*, c-22(1):67, January 1973.
- [8] M. Ichino. General metrics for mixed features, the cartesian space theory for pattern recognition. In *IEEE International Conference on Systems, Man And Cybernetics*, 1988.
- [9] M. Ichino. Generalized minkowsky metrics for mixed feature type data analysis. *IEEE Transactions on Systems, Man And Cybernetics*, 24(4), 1994.
- [10] T. Kanade. *Picture processing system by computer complex and recognition of human faces*. PhD thesis, Kyoto University, Department of Information Science, November 1973.
- [11] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. In *IEEE Proceedings of the International Conference on Computer Vision*, pages 259–268, London, June 1987.
- [12] P. Nagabushan. *An efficient method for classifying remotely sensed data, incorporating dimensionality reduction*. PhD thesis, Mysore University, India, 1988.
- [13] E. Person and K. Fu. Shape discrimination using fourier descriptors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(3):388–397, May 1986.
- [14] L. H. Staib and J. S. Duncan. Boundary finding with parametrically deformable models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(11):1061–1075, November 1992.
- [15] M. Turk. *Interactive-time vision: face recognition as a visual behavior*. PhD thesis, Massachusetts Institute of Technology, September 1991.
- [16] A. Yuille, D. Cohen, and P. Hallinan. Facial feature extraction by deformable templates. Technical Report 88-2, Harvard Robotics Laboratory, 1988.

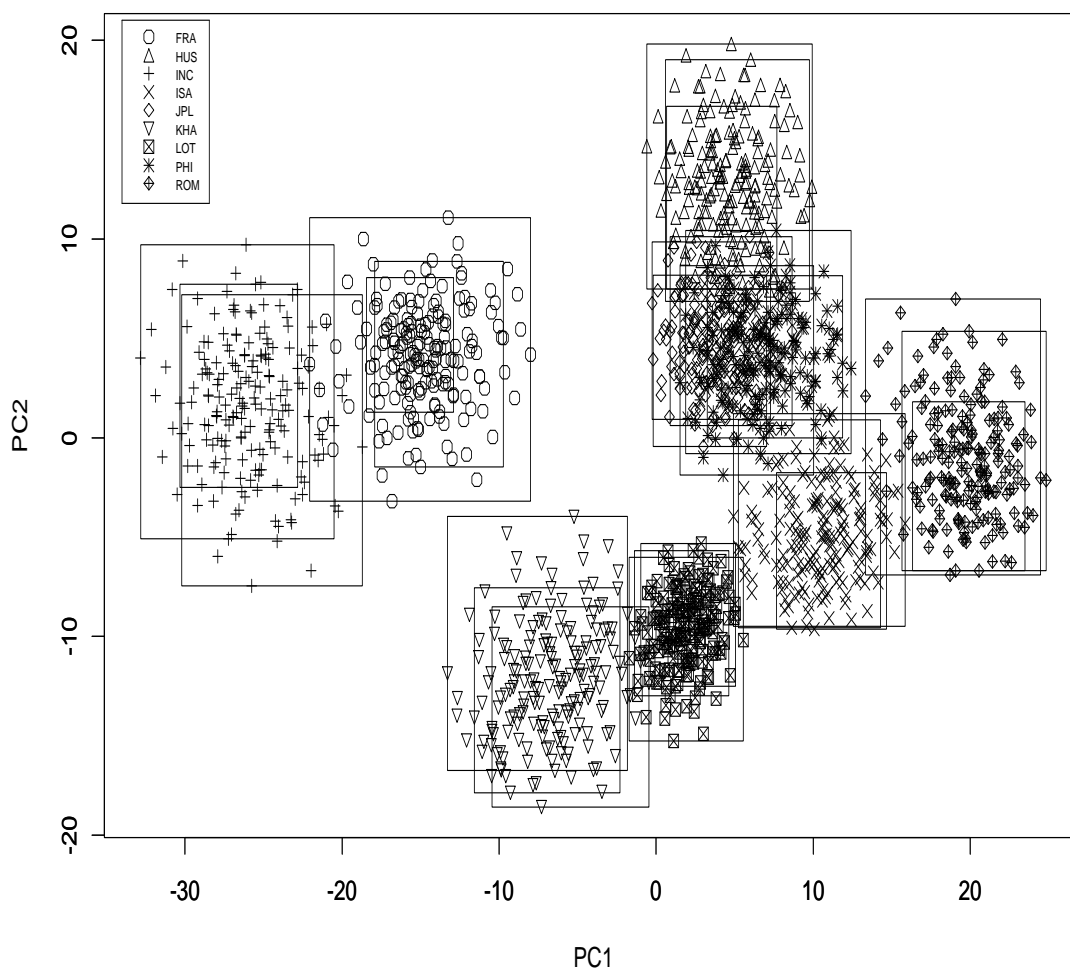


FIG. 10 - Visualisation des rectangles et des sommets associés aux objets visages dans le plan factoriel défini par les deux premières composantes principales (PC1 et PC2).

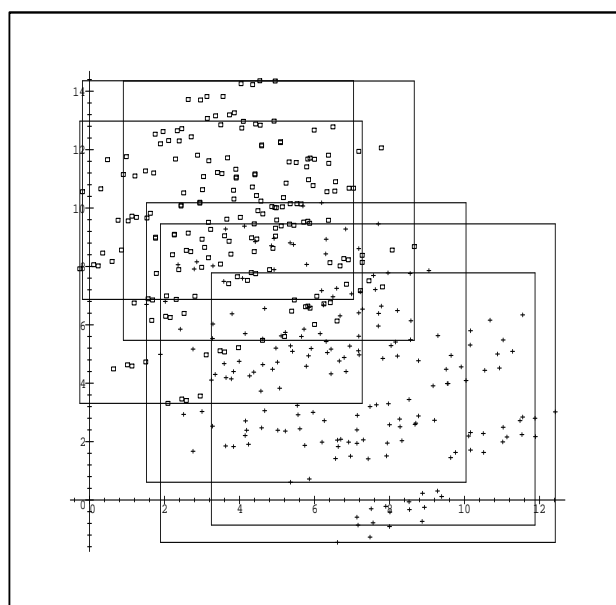


FIG. 11 - Visualisation des rectangles et des sommets associés aux objets visages JPL et PHI dans le plan factoriel défini par la première et la troisième composantes principales (PC1 et PC3).